

Statistická analýza dat II (PSY252) – seminární skupina A

Úkol 4

Zpracovaly: Denisa Lakomá, 414925

Kristýna Cejthamrová, 362244

Pro naši práci jsme využily datový soubor ESS5e03.sav, který sestával z 52458 respondentů. Z tohoto počtu bylo 23782 mužů (45,3 %) a 28655 žen (54,6 %). Věkové rozmezí se pohybovalo od 15 do 102 let ($M=48,5$; $SD=18,79$). Pro analýzu rozptylu vzorek čítal 43267, protože někteří z respondentů neuvodili odpověď na položky použité v naší hypotéze, proto byli z analýzy vyřazeni.

1) One-way ANOVA

Jako intervalovou závislou proměnnou pro analýzu rozptylu jsme zvolily celkový počet odpracovaných hodin týdně včetně přesčasů (*Total hours normally worked per week in main job overtime included*), popisné statistiky jsou uvedeny v Tabulce 1. Kategorickou nezávislou proměnnou pro nás pak byla průměrná doba strávená sledováním televize v pracovní dny (*TV watching, total time on average weekday*), popisné statistiky pro druhou proměnnou jsou obsaženy v Tabulce 2. Tyto proměnné jsme vybraly proto, že lze mezi nimi očekávat určitý vztah, a můžeme si dopředu stanovit hypotézu: **Jedinci, kteří mají čas více sledovat televizi, tráví méně času v zaměstnání.**

Zvolené proměnné splňují předpoklady pro použití one-way ANOVA. Data byla získána na základě nezávislého pozorování. Rozložení proměnných je normální, u proměnné „sledování televize“ poslední kategorie (více než 3 hodiny) narušuje normalitu, pracujeme však s dostatečně velkým vzorkem a v každé sledované skupině je natolik velký počet respondentů, že by to výsledky naší analýzy nemělo ovlivnit. Levenův test homogenity rozptylů vyšel signifikantní, ale protože nejvyšší hodnota rozptylu se od nejnižší výrazně nelišila ($s^2_{\min}=161$; $s^2_{\max}=225$), předpokládáme, že podmínka homogenity rozptylů je také splněna.

Analýza rozptylu vyšla signifikantní, kdy $F(7,43259)=5,867$; $p<0,05$; $\eta^2=0,001$. Proto mezi skupinami v rámci kategorické proměnné předpokládáme rozdíl, který budeme

testovat plánovaným kontrastem. Náš model však vysvětluje pouze 0,01 % vysvětleného rozptylu.

Naše hypotéza pro plánované kontrasty předpokládá, že lidé, kteří se dívají na televizi více, než dvě hodiny denně, mají průměrně nižší počet strávených hodin v práci než ti, kteří se dívají na televizi méně než 2 hodiny nebo vůbec. Popisné statistiky pro jednotlivé skupiny jsou uvedeny v Tabulce 1, kategorie jsme tedy rozdělili na dvě skupiny: 0 – 2 hodiny denně a 2 a více hodin denně. Kontrast mezi zvolenými skupinami nebyl signifikantním, $t(43259)=-0,241$, $p=0,81$, $r=0,5$, proto nemůžeme říct, že mezi námi zvolenými skupinami byl nalezen očekávaný rozdíl.

Tabulka 1. Popisné statistiky nezávislé proměnné sledování televize

	četnosti	%
vůbec	2166	4,1
méně než 0,5 hodiny	2586	4,9
0,5 - 1 hodin	6659	12,7
1 - 1,5 hodin	6733	12,8
1,5 - 2 hodiny	8518	16,2
2 - 2,5 hodiny	6488	12,4
2,5 - 3 hodiny	6569	12,5
více než 3 hodiny	12610	24,0
chybějící hodnoty	129	0,2
celkem	52458	100,0

Tabulka 2. Popisné statistiky pro závislou proměnnou počet hodin strávený průměrně v práci

	N	Min	Max	Průměr (M)	SD
čas strávený v práci v hodinách	43351	0	168	40,54	13,36

*respondenti s odpovědí nad 120 (N=17) byli z analýzy vyřazeni

2) Faktoriální ANOVA

V rámci faktoriální analýzy rozptylu jsme pracovaly se stejnými proměnnými jako v předešlém případě s tím, že jsme navíc do modelu jako další nezávislou kategorickou proměnnou zahrnuly to, jak často se jedinec zapojuje do společenských aktivit v porovnání s jinými lidmi ze stejné věkové kategorie (*Take part in social*

activities compared to others of same age) (Tabulka 3). Předpokládáme, že **lidé, kteří tráví více času společenskými aktivitami a více se dívají na televizi, tráví méně času v práci.**

Tabulka 3. Popisné statistiky nezávislé proměnné účast na sociálních aktivitách v porovnání s vrstevníky

	četnosti	%
mnohem méně než ostatní	5797	11,1
méně než ostatní	13243	25,2
stejně	24003	45,8
více než ostatní	6776	12,9
mnohem více než ostatní	1519	2,9
chybějící hodnoty	1120	2,1
celkem	52458	100

Zvolený model splňuje předpoklady pro použití analýzy rozptylu. Vybrané proměnné mají normální rozložení, Levenův test vyšel signifikantní ($p < 0,05$), nicméně stejně jako v předešlém případě pracujeme s dostatečně velkým vzorkem a ze srovnání rozptylů je patrné, že homogenita rozptylu není porušena. Pro každou kombinaci faktorů je navíc zastoupen dostatečný počet případů (nejmenší skupina čítá 54 jedinců).

Výsledná interakce obou nezávislých proměnných byla shledána signifikantní, $F(28,42504)=1,683$; $p=0,01$; $\eta^2=0,001$. Změna velikosti účinku je však oproti případu, kdy jsme nezapočítaly další kategorickou proměnnou, minimální. Z grafu je patrné, že lidé, kteří vstupují do sociálních interakcí méně často, tráví v práci více hodin, počet hodin v práci pak zároveň klesá s rostoucím počtem hodin strávených u televize. Lidé, kteří často vstupují do sociálních interakcí, tráví naopak méně času v práci, s rostoucím počtem hodin, které se respondenti dívají na televizi, však čas strávený s přáteli a v práci zůstává téměř stejný.